# Big Data Networking at Fermilab

*Phil Demar, David Dykstra, Gabriele Garzoglio, Parag Mhashilkar, Anupam Rajendran, Wenji Wu*

## *Spectrum of Support for Data Movement and Analysis in Big Data Science*

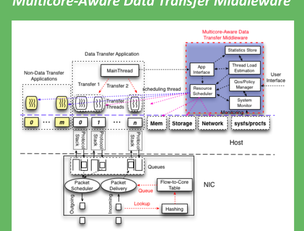
## Performance Optimization and Tools
## G-Netmon & MDTM

### *Multicore-Aware Data Transfer Middleware (MDTM) - Harnessing Multicore Parallelism to Scale Data Movement Toolkits*

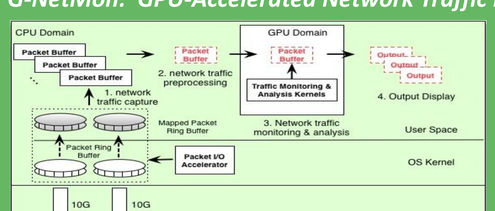
MDTM is a user-space resource scheduler that harness multicore parallelism to scale data movement toolkits at multicore systems.

- Data Transfer-Centric Scheduling and Resource Management Capabilities.
- NUMA Topology-Aware Scheduler
- Supporting Core Affinity on Networking Processing
- Supporting the QoS Mechanism to allow Differentiated Data Transfer

*Multicore-Aware Data Transfer Middleware*




### *G-NetMon: GPU-Accelerated Network Traffic Monitoring & Analysis Architecture*




- A combination of multicore and manycore technologies
- Real-time Network Traffic Monitoring & Analysis
- Handling 10,000,000+ pps capabilities

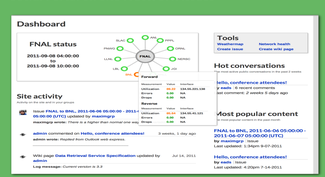
## Network Management and Control
## E-Center & ESCPS

*Network path performance weather map*

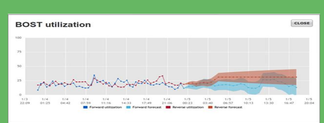

### *E-Center*

- Provides end-to-end and hop-by-hop network path data extending across multiple network domains.
- Portal for end users to capture and discuss network path-related issues and interpretation of
- Perfsonar-collected data with experts
- Traffic forecasting capability for user-specified network paths
- ***https://ecenter.fnal.gov***

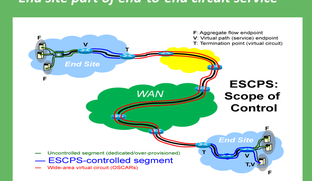
*Social Portal on Site Network Issues*




*Forecasting network traffic conditions*



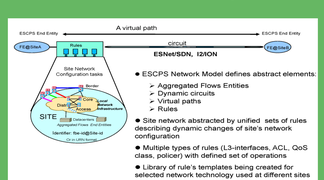

*End site part of end-to-end circuit service*




### *End Site Control Plane System*

- Site integration with wide-area network circuit service
- Local network configured for circuit-directed traffic

*Configuration rules define routing modifications*




## High Performance Data Program (HTDP)

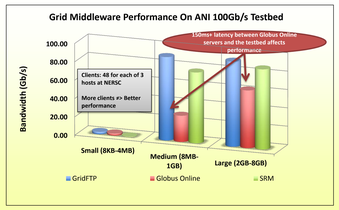
**Mission**

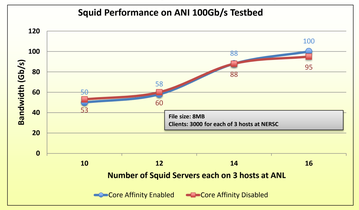
Mission of the High Throughput Data Program (HTDP) at Fermilab is to prepare the Laboratory and its stakeholders for 100GE infrastructure.

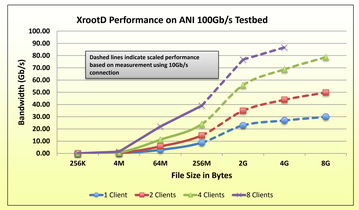
**Focus**

Test the performance and identify gaps in the Grid Middleware used by Fermilab stakeholders on the 100GE testbed operated by ESnet's Advanced Networking Initiative (ANI)

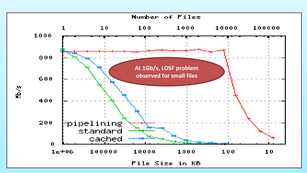
**Performance Analysis on ANI 100Gb/s Testbed**

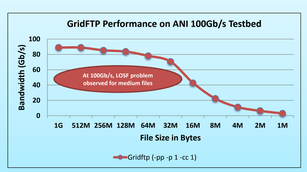



**Lots Of Small Files (LOSF) Problem on 1Gb/s v/s 100Gb/s**
At 100Gb/s, Medium Is The New Small



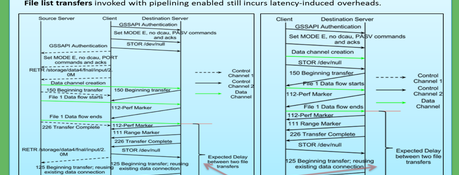




Reference: John Bresnahan, Michael Link, Rajkumar Kettimuthu, Dan Fraser, Ian Foster; "GridFTP Pipelining", In the proceedings of Teragrid 2007 Conference, Madison, WI, USA

Acknowledgement: This material is based upon work supported by the National Science Foundation under Grant No. 1007115, ExTENCI: Extending Science Through Enhanced National Cyberinfrastructure. This research used resources of the ESnet Testbed, which is supported by the Office of Science of the U.S. Department of Energy under contract DE-AC02-05CH11231

**Shortcomings of the GridFTP Pipelining Mechanism**

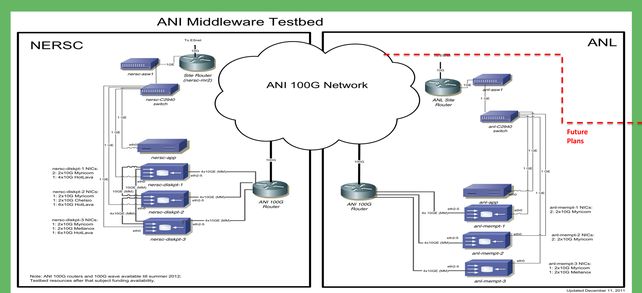
The implementation of pipelining in GridFTP works only for **third-party whole directory transfers**. **File list transfers** invoked with pipelining enabled still incurs latency-induced overheads.

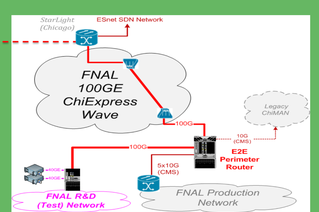



## 100GE Test and Evaluation Environment

**ESnet Advanced Network Initiative (ANI) 100GE Testbed**

- Six hosts, three each at NERSC & ANL (RTT: 54ms). Each host has four 10 GE NICs.
- Test Datasets: Files from 8KB to 8GB increasing in power of 2, split into three sets, small, medium and large. Each dataset contained files of different sizes.




**100GE at Fermilab**

- Dedicated 100 GE wave into Fermilab (ESnet ChiExpress MAN)
- 100 GE connection will be used for R&D, connecting Fermilab to ESnet ANI testbed.
- Local testbed at Fermilab will include 10GE & 40GE host systems.

**HTDP Future Plans**

- Facilitate performance testing before moving 100 Gb/s to production -
  - Redo ANI tests using local testbed
  - Test the performance of other Grid Middleware tools like dCache with NFS4, CVMFS, IRODS, etc., based on the stakeholders needs